Transdigital
revista científica

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

Artículo de Investigación

Análisis con minería de datos del Marco Curricular y el Plan de Estudios 2022 de la Educación Básica Mexicana

Analysis with data mining of the Curricular Framework and the 2022 curriculum of Mexican Basic Education

Ulises Alejandro Duarte Velázquez
Universidad Vasco de Quiroga, México. aduartev@uvaq.edu.mx
ORCID: https://orcid.org/0000-0002-6527-9845

Sección: **Artículo de investigación**
Fecha de recepción: **11/06/2022** | Fecha de aceptación: **29/06/2022**

Referencia del artículo en estilo APA 7ª. edición:

Duarte Velázquez, U. A. (2022). Análisis con minería de datos del Marco Curricular y el Plan de Estudios 2022 de la Educación Básica Mexicana. *Transdigital*, 3(6), 1–34. https://doi.org/10.56162/transdigital122

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

Resumen

La educación debería estar pautada por una ruta clara, de acuerdo con los Objetivos de Desarrollo Sostenible 2030. Esta ruta tendría el objetivo de garantizar una educación de calidad y el bienestar de la sociedad, por lo que México tiene la necesidad de establecer un nuevo marco curricular a favor de niñas, niños y adolescentes que viven situaciones de desigualdad en la educación básica, en especial después de la pandemia COVID-19. El propósito de este estudio fue analizar el documento de propuesta del Marco Curricular y Plan de Estudios 2022 para encontrar aquellos elementos centrales que lo componen y así establecer una comparación con sus predecesores, Aprendizajes Clave para la Educacion Integral 2017 y Plan de Estudios 2011 Educación Básica. La comparación fue realizada con minería de texto usando *R Studio* con diversos paquetes como *Quanteda*, *Tidyverse*, *word2vec*. Se construyó el *corpus* con los tres modelos educativos más recientes de México. El texto se preprocesó para ejecutar diversos análisis con métodos supervisados y no supervisados. Los resultados mediante el procesamiento de texto arrojan marcadas diferencias semánticas y estilo métricas: el 2011 se centra en la disciplina; el 2017 está marcado por competencias y la eficiencia; finalmente, el 2022 subraya el desarrollo de la comunidad buscando trascender la disciplina, con un impulso mayor de la inclusión y el bienestar de la población en desigualdad.

Palabras clave: Modelo educativo; minería de texto; sistema educativo; reforma educativa; lenguaje de programación.

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

Abstract

Education should be guided by a clear path, in accordance with the Sustainable Development Goals 2030. This path would have the objective of guaranteeing quality education and the well-being of society, for which Mexico has the need to establish a new framework curriculum in favor of children and adolescents who experience situations of inequality in basic education, especially after the COVID-19 pandemic. The purpose of this study was to analyze the proposal document of the Curricular Framework and Study Plan 2022 to find those central elements that compose it and thus establish a comparison with its predecessors, Key Learning for Integral Education 2017 and Study Plan 2011 Basic Education. The comparison was made with text mining using *R Studio* with various packages such as *Quanteda*, *Tidyverse*, *word2vec*. The corpus was built with the three most recent educational models in Mexico. The text was preprocessed to run various analyzes with supervised and unsupervised methods. The results through text processing show marked differences in semantics and metric style: 2011 focuses on the discipline; 2017 is marked by skills and efficiency; finally, 2022 underlines the development of the community seeking to transcend the discipline, with a greater drive for the inclusion and well-being of the population in inequality.

Keywords: Educational model; text mining; educational system; educational reform; programming language.

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

1. Introducción

En la actualidad, la información crece a ritmos acelerados. Diversas estimaciones indican que el 80% de esta información son textos (Anawis, 2014), por lo que analizar grandes cantidades de información ha llevado a la minería de datos a ser una herramienta útil para acotarla. La minería de textos es un conjunto de técnicas estadísticas desarrolladas para analizar información escrita (Stein-Sparvieri, 2010). Tiene diversas aplicaciones. Desde lo más simple, que es el conteo de palabras, hasta técnicas complejas, como son la vinculación entre palabras, creación de filtros para correos *spam*, análisis de sentimiento, tendencias, etcétera. La minería de textos, por ende, tiene su origen en la inteligencia artificial, la estadística y la gestión de bases de datos, cuyo propósito es procesar textos con varios algoritmos estadísticos (Zanini & Dhawan, 2015).

Este proceso se realiza comunmente a través del procesamiento del lenguaje natural (PNL, por sus siglas en inglés), que es la forma en que una máquina se comunica con los humanos mediante la inteligencia artificial, por lo que se requiere una computadora en interacción con un humano (Maulud et al., 2021). La minería de texto se ha utilizado en diversas áreas de conocimiento, como la política, la psicología y la medicina. En esta última tenemos el ejemplo del análisis del compartimiento de pacientes en foros de bebidas alcohólicas, mediante un marco semántico basando en el PNL y *word2vec* para crear un modelados de temas con Distribución Latente de Dirichlet (LDA por su siglas en inglés) mediante la extracción de palabras y su similitud de coseno (Jelodar et al., 2020).

Las primeras aplicaciones de la minería de textos datan de 1980 (Zanini & Dhawan, 2015). En la investigación aplicada se usa, por ejemplo, en la educación. El objetivo es minar datos educativos como disciplina emergente para comprender mejor los procesos de enseñanza aprendizaje en temas como deserción académica, bajo rendimiento educativo o

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

éxito educativo (Aleem & Gore, 2020). Es por ello que el presente artículo busca llevar el procesamiento del lenguaje natural con la minería de texto a comprender los documentos normativos de la educación mexicana con una comparación basada en minería de datos educativos.

Aunando lo anterior, en México y el mundo se vive la postpandemia, ocasionada por el virus Sars-Cov2 o COVID. Las escuelas migraron de un modelo presencial a un modelo de educación a distancia para sortear el problema (König et al., 2020; Moawad, 2020). Esto, sin duda, agudizó diversos problemas educativos, afectando a más del 90% de la población estudiantil mundial (Chandra, 2020) con una perdida en el logro de los aprendizajes específica o general de conocimientos y habilidades o reversiones en el progreso académico (Rieble-Aubourg & Viteri, 2020). La perdida estimada es de 0.6 años de escolaridad (Hevia et al., 2022), afectando de manera considerable a las personas en desventaja social. Específicamente, en los países más pobres como México, más de 700 mil estudiantes entre los 3 y los 29 años no terminaron el ciclo escolar 2019-2020. En el ciclo escolar 2020-2021, 5.2 millones no se matricularon debido a la pandemia (INEGI, 2021). Esto implica un aumento en la deserción.

Lo anterior, frente al marco de una Nueva Escuela Mexicana, invita a las personas interesadas en la educación de excelencia o de calidad a desarrollar una mejor comprensión de los textos que rigen a la educación. Se buscan nuevas formas de comprensión apoyadas de la inteligencia artificial para potenciar la compresión humana. Para ello, los patrones pedagógicos que se establecen en los diferentes documentos normativos ayudarán a descubrir la ideología más amplia de la que forma parte un modelo educativo.

El Modelo Educativo 2022, aún en desarrollo, inició su proceso de construcción en julio de 2021 con 260 personas de 24 instituciones y dependencias gubernamentales. Ellos

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

generaron alrededor de 1,423 reuniones de trabajo (virtuales y presenciales), lo que dio como resultado el Marco Curricular, el Plan de Estudios y 15 Programas de Estudio (SEP, 2022a). Cabe mencionar que, aunque existen análisis del modelo educativo del 2017 (Flores, 2017; Martínez Iñiguez et al., 2020), el Nuevo Modelo Educativo 2022, que regirá la Escuela Mexicana, no cuenta con análisis profundos.

Solo existe un proceso de análisis por parte del Gobierno de México, dirigido por Marx Arriaga Navarro, Director General de Materiales Educativos de la Secretaría de Educación Pública, que consistió en realizar asambleas con docentes de forma presencial y con trasmisión por redes sociales. Asimismo, se llenaron formularios de consulta por parte de docentes entre el 31 de enero al 25 de marzo de 2022 (SEP, 2022b).

La educación tiene una tendencia a la adaptación del sistema educativo a la globalización económica, término que se ha convertido en movilizador del esfuerzo social y económico. Pero ¿dónde comenzaron las reformas educativas? Estas iniciaron de manera más clara cuando los países buscaban una visión de cómo deberían ser la sociedad después de una segunda postguerra mundial (Cowen, 2018). Sin duda alguna, las nuevas reformas educativas que surjan deberán estar marcadas por la visión de una postpandemia, para que no tome por sorpresa a los sistemas educativos con situaciones similares al COVID-19 y evitar pérdidas de aprendizaje, que afectan a la población más vulnerable.

En México la igualdad de oportunidades educativas conlleva a desarrollar cambios en los modelos educativos buscando la respuesta a ¿cuál es la mejor propuesta curricular para este segunda década del siglo XXI? Esta respuesta debería estar marcada por los Objetivos del Desarrollo Sostenible 2030. En específico, en el objetivo número cuatro que es “garantizar una educación inclusiva, equitativa y de calidad y promover oportunidades de aprendizaje

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

durante toda la vida para todos" (ONU, 2013), la cual busca ciudadanos democráticos y no solamente fuerza laboral alfabetizada.

Aunque también podemos encontrarnos con sistemas educativos con resultados excelentes (OECD, 2018) como China, que es el país con mejores resultados en PISA, con un sistema educativo que se ha caracterizado por el adoctrinamiento político y que, a su vez, tuvo una reforma educativa en 2020 (Lo & Hung, 2022). Este sistema se caracteruza por el aprendizaje entre pares maestro-maestro, mediante la colaboración, para que los docentes de menos experiencia aprendan de los de mayor experiencia. Asimismo, es necesario el apoyo para escuelas desfavorecidas, apoyo para favorecer las habilidades de enseñanza y liderazgo mediante programas de desarrollo profesional, promoción de ambientes de aprendizaje con relaciones positivas (Lewis, 2020).

Otro país reconocido por su buen sistema educativo es Singapur. Estudios de la Organización para la Cooperación y el Desarrollo Económico (OCDE) mencionan que los estudiantes de Singapur se preocupan por las malas calificaciones más que el promedio del resto de países pertenecientes a la OCDE (Ng, 2020). Además, "Since its inception as a nation state in 1965, the ideology of meritocracy has been enshrined as a key principle of governance and educational distribution" (Lim & Tan, 2018), donde, para que la meritocracia funcione, requiere un principio de no discriminación. Así las recompensas se establecen en función del talento y calificación de las personas. Además de ser un sistema educativo donde el trabajo por proyectos creados por los alumnos termina cuando estos comparten sus productos con la comunidad. Otro factor de éxito de Singapur es que reclutan y capacitan a buenos docentes, con inversión en su formación, impulso al liderazgo escolar, altos estándares para los docentes futuros así como evaluaciones anuales (Lewis, 2020).

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

Finalmente, es obligación hablar de Finlandia como un sistema educativo de excelencia, cuando la intención de México es desarrollar un Marco Curricular 2022 que rija en los próximos años. Algo que destaca de ese país es que dicha educación se basa en la confianza y la responsabilidad, buscando asegurar la calidad con apoyo y no con control administrativo, con una escuela como comunidad de aprendizaje (Finnish National Agency for Education, 2016), evaluación de estudiantes con base a su progreso individual y no con indicadores estadísticos (Lewis, 2020). De igual manera, México busca como principio, desarrollar el aprendizaje desde la comunidad. El sistema educativo de dicho país es descentralizado por lo cual la autonomía curricular es muy fomentada para el docente, escuela y municipio, la cual tiene pilares claros en un curriculum basado en estándares, docentes altamente capacitados, además de una educación basada en la equidad e inclusión (Andere, 2014).

Otro tema que está ligando a la Nueva Escuela Mexicana es que el gobierno actual de Andrés Manuel López Obrador ha hecho aseveraciones sobre la privatización de la educación (Jiménez, 2021; Vallejo, 2021), donde la "privatización de la educación es una política que corre el riesgo de socavar la equidad y cuyos supuestos beneficios, en términos de eficiencia o mejoras de calidad, no se ha probado empírica y rigurosamente a nivel mundial" (Verger et al., 2016). También existe una idea sobre el neoliberalismo. A dicha corriente, el Gobierno de México lo considera como un régimen de corrupción (López Obrador, 2020). Tomando en cuenta esta visión política, se puede decir de manera breve que el neoliberalismo se desarrolló durante la década de 1950, pero comenzó a tener relevancia en las políticas públicas en la década de 1970 (Klees, 2008), las cuales se basa en la "privatización de activos tradicionalmente públicos" (Fitz & Hafid, 2016).

Acorde con lo expuesto, el presente estudio se orientó en torno a los siguientes propósitos: 1) Obtener el modelado de temas Marco Curricular y Plan de Estudios 2022 de la

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

Educación Básica Mexicana 2022 (MCyPE 2022), Modelo 2017 y Modelo 2011; 2) Reconocer los términos más importantes de Marco MCyPE 2022, Modelo 2017 y Modelo 2011 con *TF-IDF*, *Wordfish* y *Word embbeding*.

2. Método de investigación

El tema consiste en hacer análisis cuantitativo de texto con el fin de comparar las propuestas educativas de los últimos 15 años en México, desde sus documentos rectores, que son el MCyPE 2022 al cual nombraremos Modelo 2022, el Modelo Educativo 2011 y el Modelo Educativo 2017 mejor conocido como Aprendizajes Clave para la Educación Integral. De esta manera se pretende ampliar el conocimiento que se puede genera en conjunto con la lectura reflexiva y poner en el contexto de la educación en México las diferencias y similitudes que existen entres los últimos tres modelos educativos. Quantitative text analysis o análisis cuantitativo de textos (QTA), permite realizar procesos de análisis sistemático de colecciones de texto de manera automática, lo cual ayuda a obtener diversas manifestaciones de la estructura de estos. Sin embargo, los métodos automatizados nunca remplazarán la lectura cuidadosa, más bien amplían la comprensión de análisis reflexivo (Grimmer & Stewart, 2013).

De manera sintética, cuando se realiza minería de texto se recolectan los documentos a analizar construyendo un *corpus*, los cuales tendrán que ayudar a cumplir los objetivos de la investigación. Así se toma la decisión de hacer el proceso analítico mediante *wordscores*, con un método de aprendizaje supervisado, que compara las frecuencias relativas de palabras utilizadas en textos de referencia. O mediante *wordfish*, basada en el aprendizaje no supervisado, que analiza frecuencias de palabras mediante la distribución de Poisson y el uso de un algoritmo de maximización de probabilidad (Leonisio & Strijbis, 2012). La clasificación de documentos se puede realizar cuando se conocen la categorías, como lo es

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

el caso de uso de diccionarios, donde hay ejemplos claros que se usan para realizar análisis de sentimiento (Silge & Robinson, 2017).

2.1. Procedimiento

De acuerdo con los propósitos de la investigación, el estudio se desarrolló mediante las siguientes fases:

Fase 1. Limpieza de los documentos.

Fase 2. Cálculo de estilometría.

Fase 2. Cálculo de *TF-IDF*.

Fase 3. Modelado de temas con *Asignación Latente* de *Dirichlet*.

Fase 4. *Wordfish*.

Fase 5. *Glove*.

Los cálculos se realizaron utilizando una computadora con las siguientes características: Intel(R) Core(TM) i5-7400 CPU de 3.00GHz, 8 GB de memora DDR3. Ejecutando lenguaje de programación en R Estudio versión 1.4.1717.

2.1.1. Limpieza de los documentos

Para la limpieza y preprocesamiento del *corpus*, se descargaron los documentos de las páginas oficiales. Posteriormente, mediante una limpieza manual, se quitaron imágenes, citas al pie, para dejar de manera exclusiva el texto. En *R Studio* se trabajaron con las siguientes paqueterías con el siguiente lenguaje de programación para extraer el texto: *tabulizer*. *tidyverse*, *pdftools*, *tidytext*

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

```
d2011 <- extract_text("2011.pdf")

d2017 <- extract_text("2017.pdf")

d2022 <- extract_text("2022.pdf")
```

Después, para limpiar de mejor manera el texto, se procedió con el siguiente lenguaje de programación en *R Studio*, con cada uno de los documentos del corpus, con la intención de *tokenizar* los términos. Además de eliminar las palabras vacías en español (dichas palabras son las más comunes en un idioma) (Cheng et al., 2020), remover guiones, números, puntuaciones y dejar solamente palabras con más de dos caracteres o letras, y y convertir el término Sars-CoV2 en COVID ya que se podría perder en el análisis, por ser un bigrama; está conversión se realizó solamente en el documento 2022.

```
c2022 <- d2022 %>%
 str_squish () %>%
 str_remove_all ("- ") %>%
 str_remove_all ("•|·|°") %>%
 str_remove_all ("[0-9]") %>%
 str_replace_all ("SARS-CoV2", "SarsCov") %>%
 str_replace_all ("¡deas", "ideas") %>%
 as_tibble () %>% # Tabla
 unnest_tokens (palabras, value) %>%
 filter(!palabras %in% tm:stopwords("es")) %>%
 filter(!palabras %in% c("ios", "baja")) %>%
 mutate(palabras = ifelse(str_detect(palabras, "sars|cov"), "covid", palabras )) %>%
```

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

```
count(palabras, sort = T) %>%
filter(nchar(palabras) > 2) %>% mutate(edicion = "2022")
```

2.1.2. Cálculo de la estilometría del corpus

Para el análisis exploratorio se ejecutó un lenguaje de programación que ayuda a obtener la estilometría de los documentos, para lo cual se cargaron las paqueterías de udpipe y tidyverse, con el modelo spanish ancora, el cual es un conjunto de capas en español con 400,000 palabras (Taulé et al., 2016) para extraer la estilometría de los documentos a analizar.

```
pacman: p_load (udpipe, tidyverse)
es <- udpipe_download_model (language = "spanish-ancora")
udmodel <- udpipe_load_model (file = es$file_model)
modelo2017 <- readtext: readtext ("2011 limpio.pdf")
m2017.txt <- modelo2017$text %>%
  str_remove_all ("- ") %>%
  str_remove_all ("•|·|°") %>%
  str_remove_all ("[0-9]")
tag <- udpipe_annotate (udmodel, x = m2017.txt)
tag <- as_tibble (tag)
```

2.1.3. Cálculo de TF-IDF

Una vez realizado el procesamiento y limpieza se comenzó a trabajar con el texto, el segundo cálculo ejecutado fue para obtener la frecuencia de término – frecuencia invertida

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

de documentos (TF-IDF) con unigramas, el cual es un método de investigación en el procesamiento del lenguaje natural "TF-IDF method determines the relative frequency of words in a specific document through an inverse proportion of the word over the entire document corpus" (Trstenjak et al., 2014), para ello se tiene la siguiente fórmula:

$$a_{ij} = tf_{ij} idf_i = tf_{ij} \text{ x } log_2 \left(\frac{N}{dfi}\right)$$

También se realizó el mismo proceso con bigramas (Anawis, 2014), el cual es un proceso que no usa solamente una palabra, sino la combinación de dos palabras en los documentos, en ambos casos se obtienen con la siguiente secuencia:

```
frecuencias <- bind_rows(c2011, c2017, c2022)

tfidf_modelos <- frecuencias %>%

bind_tf_idf(palabras, edicion, n) %>%

arrange(desc(tf_idf))
```

2.1.4. Cálculo de Modelado de temas con LDA

Para crear un modelado de temas se utilizó la *Asignación Latente* de *Dirichlet* (LDA), el cual es modelo estadístico que permite descubrir temas del corpus de documentos (Modelo 2022, 2017 y 2011), la cual ayuda a conocer la estructura semántica oculta en los textos analizados.

> "The process begins by drawing a K-dimensional Dirichlet vector θd that captures the expected proportion of words in document d that can be attributed to each topic. Then for each position (or, equivalently, for each word) in the document, indexed by n, it proceeds by sampling an indicator zd, n from a Multinomial K (θd, 1) whose positive component denotes which topic such position is associated with. The process ends by sampling the actual word indicator wd,n from a Multinomial V(Bzd,n,

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

1), where the matrix B = [β1| . . . |βK], encodes the distributions over terms in the vocabulary associated with the K topics" (Roberts et al., 2016).

El modelado de temas se vale de aprendizaje automático, donde el procesamiento del lenguaje natural aprovecha la presencia de datos en texto para que las computadoras comprendan las palabras y aprendan el patrón de los datos (Cheng et al., 2020). Para dicho proceso se utilizó la siguiente paquetería en *R estudio*: *topicmodels*, *pdftools*, *tm*, *tidytext*, *ggplot2* y *dplyr*; con la siguiente secuencia:

```
modelos <- list.files (pattern = "pdf$")

modelos <- lapply (modelos, pdf_text)

corpus_modelos <- Corpus (VectorSource (modelos))

corpus_modelos <-tm_map (corpus_modelos, content_transformer(tolower))

corpus_modelos <-tm_map (corpus_modelos, removeNumbers)

corpus_modelos <-tm_map (corpus_modelos, removeWords, stopwords("spanish"))

corpus_modelos <-tm_map (corpus_modelos, removePunctuation, preserve_intra_word_dashes = TRUE)

corpus_modelos <-tm_map (corpus_modelos, stripWhitespace)

corpus_modelos <-tm_map (corpus_modelos, removeWords, c("así", "parte", "debe", "cada", "partir", "iii", "manera","ello", "través","uso", "nnnnn"))

DTM <- DocumentTermMatrix(corpus_modelos)

Model_lda <-LDA (DTM, k=5,control=list (seed=1234))

beta_topics <-tidy (Model_lda, matrix ="beta")
```

Con esta secuencia se obtiene los valores beta (β) que sirven para calcular la probabilidad de que un término sea parte de un tópico o tema. Otro aspecto del algoritmo

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

LDA es asignar términos a los documentos de acuerdo con el modelado de temas. Para ello es necesario el valor gamma (γ), la cual: a mayor valor de gamma existirá una mayor relación entre documento y tema. Se ejecutó la siguiente secuencia:

```
tidy (DTM)%>%
filter(document==3) %>%
 arrange(desc(count))
gamma_modelos <- tidy (Model_lda, matrix ="gamma")
modelos_gamma.df <- data. frame(gamma_modelos)
modelos_gamma.df$chapter <- rep (1: dim (DTM) [1],5)
```

2.1.5. Cálculo de Wordfish

Mediante el aprendizaje no supervisado se buscó encontrar la posición política del Marco Curricular 2022 en comparación del modelo educativo 2022. Para ello se trabajó con *wordfish*, que sirve para estimar las posiciones ideológicas de los partidos políticos (Slapin & Proksch, 2008). Se utilizó dicho algoritmo para ver la posición de las palabras, para lo cual se utiliza la *tokenización* de términos y se prosiguió con el siguiente código:

Creación de corpus con Quanteda:

```
dcomp_doc <- dcomp %>%
  group_by (documento) %>%
  summarise (texto = paste (sen, collapse = " "))
corpus_modelos <- corpus (dcomp_doc, text_field = "texto")
```

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

```
names (corpus_modelos) <- c ("Modelo 2011", "Modelo 2017", "Modelo 2022")

dfmat_modelos <- corpus_modelos %>%
 tokens (remove_punct = TRUE, remove_symbols = TRUE, remove_numbers = TRUE) %>%
 dfm() %>% # Convierte a matriz documento - palabra (feature)
 dfm_remove (pattern = stopwords("es"))
```

Estimar un Wordfish:

```
tmod_wf <- textmodel_wordfish (dfmat_modelos, dir = c (3, 1))
summary(tmod_wf)
textplot_scale1d(tmod_wf)
textplot_scale1d (tmod_wf,
                  margin = "features",
                  highlighted = c( "educación", "competencias", "enseñanza", "humanista",
                   "comunidad", "pensamiento", "crítico",
                   "estándares", "aprendizaje", "aprendizajes",
                   "consejos", "neoliberal", "supervisión", "ejes",
                   "maestras", "indígena", "capital", "comunitario", "humanismo",
                   "mujeres", "inclusión", "equidad", "saberes",
                   "capital", "conocimiento", "interculturalidad", "género",
                   "territorio", "profesionalización", "formación", "humano",
                   "sociedad"))
```

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

2.1.6. Cálculo de Glove

Glove es una algoritmo de aprendizaje no supervisado, que ayuda a obtener las representaciones vectoriales de las palabras (Pennington et al., 2014), mediante la co-ocurrencia de las palabras. Es decir, palabras que aparecen juntas en una misma porción de texto. A diferencia del modelado de temas que se construyen con las frecuencias, *Glove* y *word2vec* usan un enfoque de redes neuronales para construir diversos vectores con las palabras (Clark, 2018). Para obtener la información se agregó la paquetería *text2vec*, *igraph*, *network*, *visNetwork* y *networkD3*; con el siguiente lenguaje de programación:

```
Vocabulario:
c2011_l <- list(c2011$palabras)
it = itoken (c2011_l, progressbar = T)
c2011_vocab = create_vocabulary(it)
c2011_vocab = prune_vocabulary (c2011_vocab, term_count_min = 2)

Matriz de co-ocurrencias de tokens:
c2011_vectorizer <- vocab_vectorizer(c2011_vocab)
c2011_tcm <- create_tcm (it, c2011_vectorizer, skip_grams_window = 10)

Global Vector:
c2011_glove <- GlobalVectors$new (rank = 50, x_max = 20)
c2011_wv_main <- c2011_glove$fit_transform (c2011_tcm, n_iter = 1000, convergence_tol = 0.00001)
c2011_wv_context = c2011_glove$components
```

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

Vectores de palabras:

```
c2011_word_vectors <- c2011_wv_main + t(c2011_wv_context)
```

Embeddings:

```
competencia <- c2011_word_vectors ["competencia", , drop = F]
cos_sim_comp <- sim2(x = c2011_word_vectors, y = competencia, method = "cosine", norm = "l2")
head (sort (cos_sim_comp [, 1], decreasing = T), 15)
```

Visualización de grafo:

```
comunidad <- c2022_word_vectors["comunidad", , drop = F]
cos_sim_com <- sim2(x = c2022_word_vectors, y = comunidad, method = "cosine", norm = "l2")
com_probs <- head (sort (cos_sim_com [, 1], decreasing = T), 50)
pal <- names(com_probs)

com_probs_t <- tibble (to = pal, prob = com_probs) %>%
   filter (prob < 0.99) %>%
   mutate (from = "comunidad") %>%
   select (from, to, prob)
```

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

3. Resultados

3.1. Estilometría de los documentos

Como se puede observar en la Tabla 1, el documento con más palabras es la propuesta 2017 con una densidad de vocabulario de 0.10. Aunque la propuesta 2022 tiene menos palabras, este posee más palabras únicas arrojando una densidad de vocabulario mayor 0.12. En cuanto a estilometría, se encontró que el Modelo 2011 posee 6351 sustantivos, 2478 adjetivos y 2173 verbos; el Modelo 2017 posee 9610 sustantivos, 3875 adjetivos y 3180 verbos; finalmente, el Modelo 2022 asume 8735 sustantivos, 3335 adjetivos y 2546 verbos. Para ello, la diversidad léxica de un documento se calcula en relación con las palabras únicas entre el total de palabras que tiene dicho documento (Maamuujav, 2021).

Tabla 1

Estilometría de los documentos

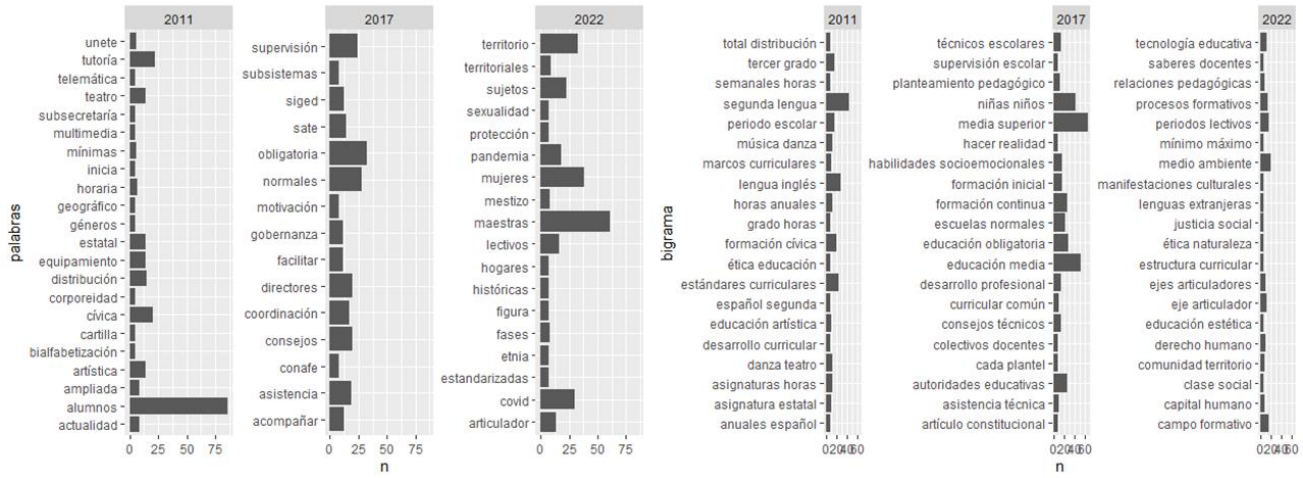
	Modelo 2011	Modelo 2017	Modelo 2022
Palabras	28056	40911	36291
Palabras únicas	3451	4332	4366
Densidad léxica	0.12	0.10	0.12
Sustantivo	6351	9610	8735
Adjetivos	2478	3875	3335
Verbos	2173	3180	2546

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

3.2. *TF-IDF* de los Modelos Educativos

Los modelos educativos tienen palabras características, como se pueden observar en la Figura 1. En el Modelo 2011, la más representativa es *alumnos*, *tutoría*, *cívica* y *artísticas*, por lo que nos podemos dar cuenta que este modelo se centra más en la disciplina (materias). En el Modelo 2017 se observan palabras características como *supeirvisión*, *normales*, *obligatoria*, *directores*, *consejos*, *coordinación*, que combinado con la lectura del mismo, se concluye como un modelo centrado en la eficiencia dirigido por las autoridades educativas. En la propuesta educativa 2022 sus palabras características son *territorio*, *mujeres*, *maestras*, *sujetos* y *COVID*-19, reflexionando con la lectura que su intención es fortalecer más la visión de comunidad y equidad de género.

Figura 1
Modelado TF-IDF



Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

Con el proceso de bigramas para el *TF-IDF* los hallazgos fueron más reveladores. En el Modelo Educativo 2011 surgieron como bigramas característicos los *estándares curriculares*, *segunda lengua* e *inglés*, mostrando con mayor claridad que dicho modelo se centraba en la especialización de las materias; el Modelo 2017 dejó clara la importancia de una *educación media superior*, la *formación continua*, la *supervisión escolar* y tomó fuerza los *consejos técnicos escolares*, manifestando con mayor claridad la tendencia de un modelo centrado en la eficiencia; mientras tanto, el Modelo Educativo 2022 tuvo elementos como *medio ambiente*, *ejes articuladores*, *proceso formativo*, *campo formativo*, *comunidad territorio*, *derecho humano*, entre otros, por lo que se devela con mayor firmeza la visión que menciona el documento una educación basada en la comunidad favoreciendo la inclusión de las personas que más padecen desigualdad.

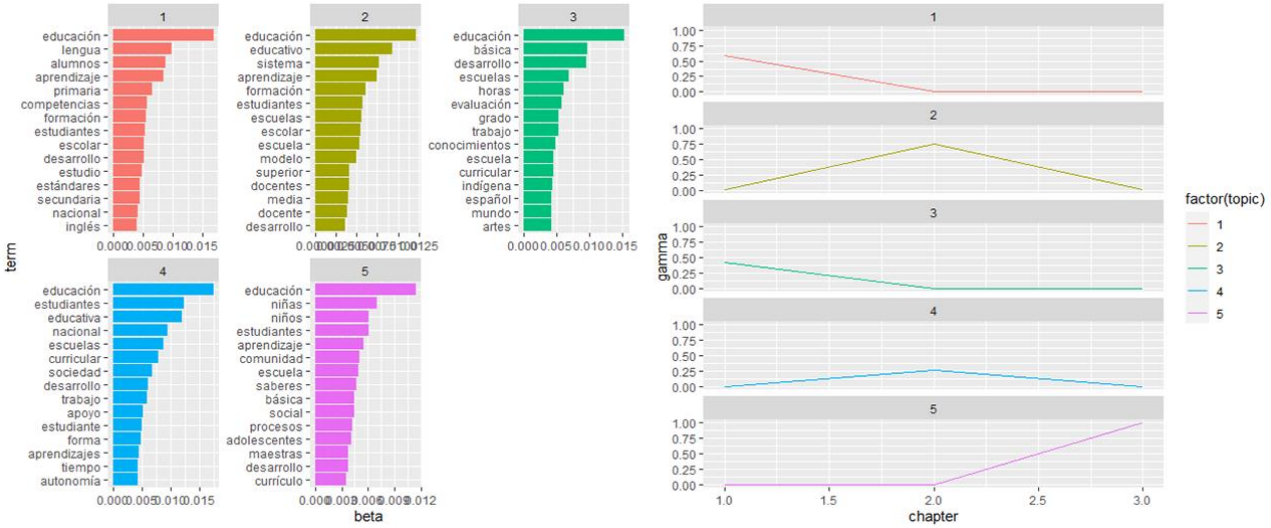
3.3. Modelado de temas

Para el modelado de temas su utilizó un k = 5, la cual arroja cinco temas y una n = 15 que extrae los quince términos más probables a aparecer juntos, que fueron calculados previamente con el valor β. Como se puede observar el la Figura 2, de acuerdo al pico del tema (factor / topic), el Modelo 2011 encaja mejor en los temas uno y tres, cuyo términos recurrentes son *lengua*, *educación*, *desarrollo*, *aprendizaje*, *competencias*, *estudiantes*, *estándares*, *inglés*, *evaluación*, *conocimientos*, *indígena*, *español*, *artes* entre otros términos. Se puede decir que el Modelo 2011 reafirmó una temática disciplinar y marcó el inicio de una visión de competencias.

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

Figura 2

Modelado de temas



En cambio, el Modelo 2017 se centró en el tema dos y cuatro, encontrando términos como *educación*, *sistema*, *aprendizaje*, *formación*, *escuelas*, *modelo*, *docente*, *desarrollo*, *estudiantes*, *escuelas*, *trabajo*, *autonomía*; con las palabras de la *TF-IDF*, se centró en un aprendizaje basado en la eficiencia y, a su vez, buscó una autonomía del docente.

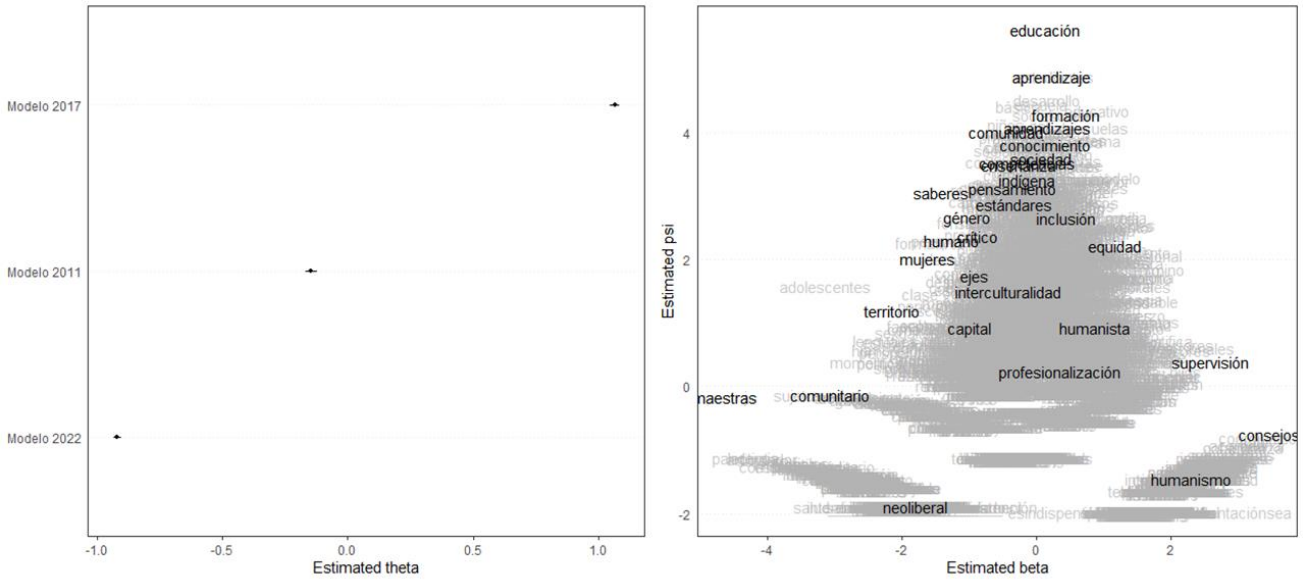
Finalmente, el Modelo 2022 encajó en el tema cinco, donde destacan términos como *educación*, *niñas*, *niños*, *estudiantes*, *aprendizaje*, *comunidad*, *escuela*, *saberes*, *social*, *procesos*, *maestras*, *desarrollo*, *currículo*, por lo que se afirma que lo dicho modelo resalta un aprendizaje basado en la comunidad, dejando clara su filosofia desde una epistemología del sur.

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

3.4. Wordfish

Como se puede observar en la Figura 3, los modelos 2022 y 2017 se polarizan; en cambio, el Modelo 2011 se quedó al centro con una ligera inclinación hacía el Modelo 2022. En cuestión de términos que están a la derecha (Modelo 2017), encontramos al *humanismo*, *supervisión*, *consejos*, *equidad* e *inclusión*; así como términos que se agrupan a la izquierda (Modelo 2022), entre ellos, *comunitario*, *maestras*, *neoliberal*, *territorio capital*, *mujeres*, *interculturalidad*, *humano*. Por lo que se puede decir que la equidad, la inclusión y el humanismo son temas recurrentes desde el 2017 y en el Modelo 2022 trae consigo una propuesta diferente en terminología, que se puede identificar con la epistemología del sur.

Figura 3
Wordfish estimado θ y β

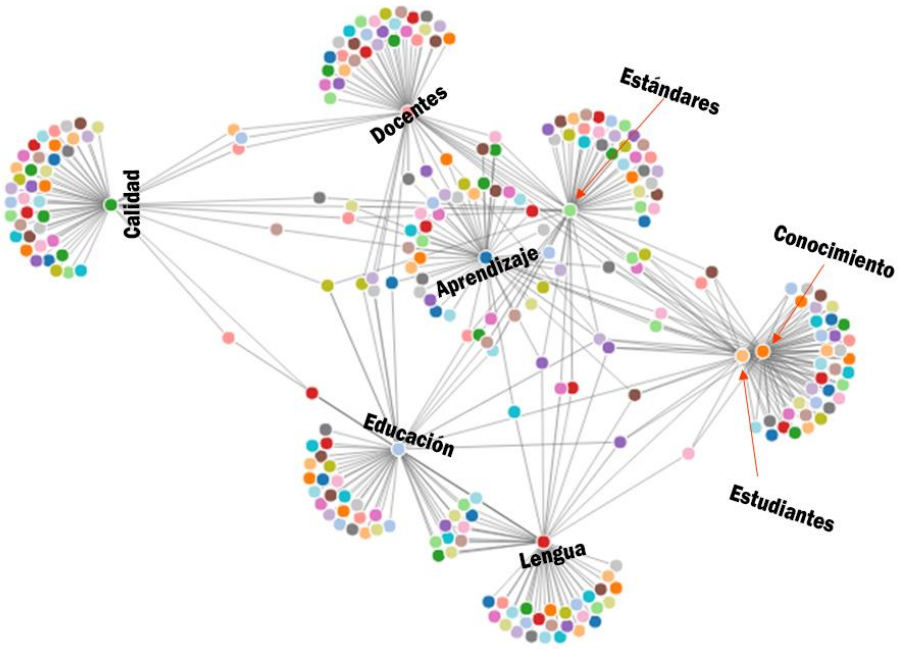


Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

3.5. Glove

Para generar el grafo (Figura 4) se seleccionaron algunas palabras características del Modelo 2011 (*calidad*, *estándares*, *lengua*, *docentes*, *conocimiento*, entre otras), con tal de tener una representación visual de términos claves y sus co-ocurrencias en el texto. Como se puede observar, la educación y el aprendizaje son clave en relación con los estándares. A su vez, los estándares están ligados al conocimiento de los estudiantes. Reafirmamos el análisis que dicha propuesta se centra en generar conocimiento eficientes por medio de los estándares para lograr una educación de calidad.

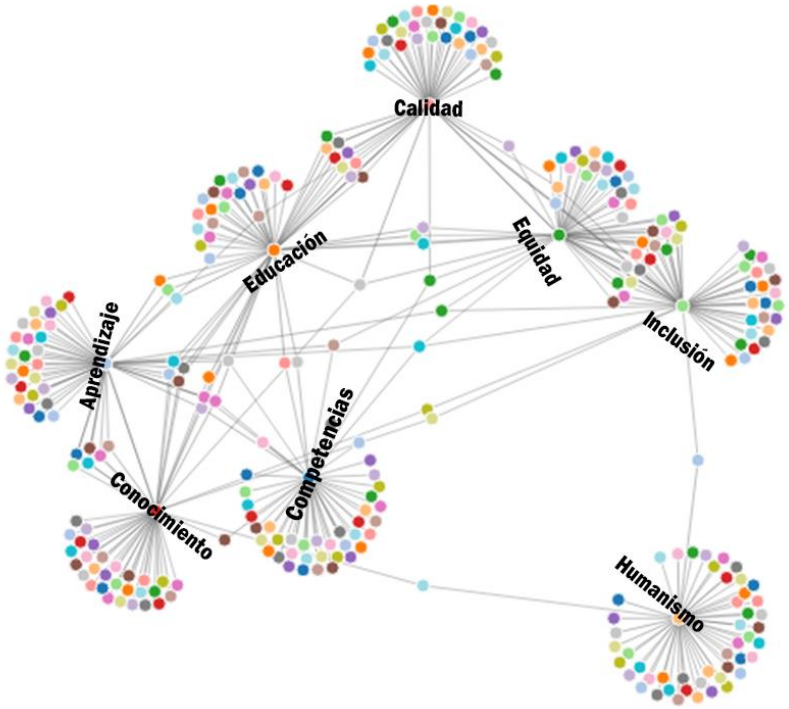
Figura 4
Grafo Modelo 2011



Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

El grafo del modelo 2017 (Figura 5) permite reconocer la importancia de ciertas palabras que son representativas como: *aprendizaje*, *conocimiento*, *calidad*, *equidad*, *inclusión* y *competencias*. Además como se puede observar que el humanismo es parte del modelo aunque tiene escasa conexión con otras categorías.

Figura 5
Grafo Modelo 2017

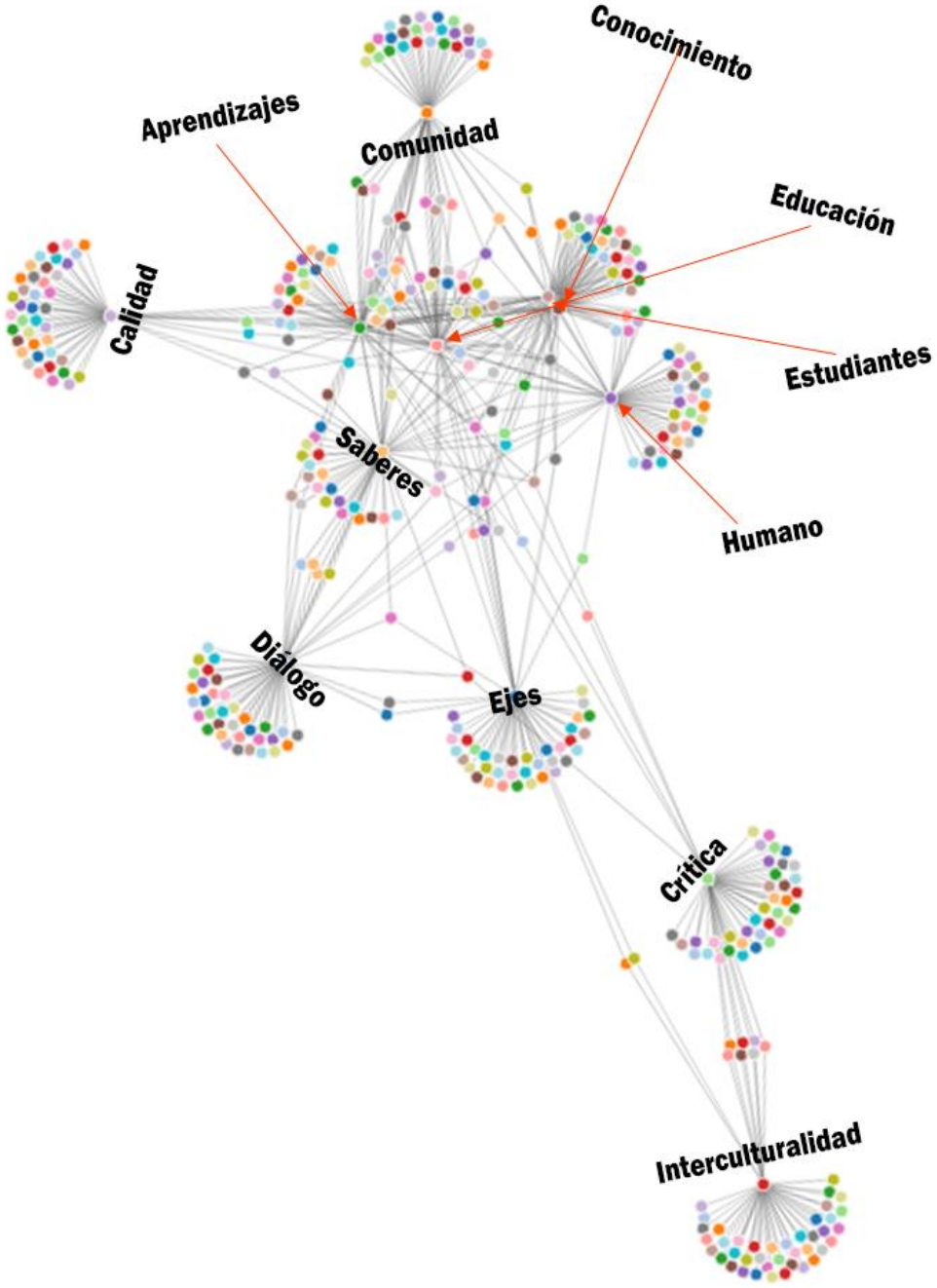


Para el grafo del modelo 2022 (Figura 6) se seleccionaron algunas palabras características del Marco Curricular 2022 (*comunidad*, *diálogo*, *eje*, *neoliberalismo*, *saberes*, *aprendizaje*, entre otras). Como se puede observar, la educación se encuentra en el centro y se conecta al diálogo por medio de los saberes; también existe un camino hacia la interculturalidad que pasa por los ejes articuladores y la crítica.

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

Figura 6

Grafo Modelo 2022



Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

4. Discusión y conclusiones

Trabajar con el *corpus* de los últimos modelos educativos, en conjunto con la propuesta 2022 para transformar el texto en número con un Aprendizaje de Máquina, potencia el análisis y clasificación temática que caracteriza a cada uno de los documentos rectores. Es importante recalcar que, la combinación de lectura reflexiva con minería de textos educativos permite tener una visión más completa, ya que ayuda a operacionalizar construcciones temáticas que no son visibles a simple vista, logrando investigar relaciones multivariadas para crear modelos de comprensión (Winne & Baker, 2013).

Sobre los propósito de la investigación, es decir, sobre obtener el modelado de temas Marco Curricular y Plan de Estudios 2022 de la Educación Básica Mexicana 2022 (MCyPE 2022), Modelo 2017 y Modelo 2011, el preprocesamiento del corpus de documentos, la *tokenización* y exclusión de palabras vacías, permitió alcanzar de manera clara cinco tópicos de manera probabilística que mejoran la interpretabilidad del tema general (Sakurai, 2012).

En cuanto al propósito dos, reconocer los términos más importantes de Marco MCyPE 2022, Modelo 2017 y Modelo 2011 con *TF-IDF*, *Wordfish* y *Word embbeding*, fue posible entender con mayor claridad los términos característicos de cada Modelo Educativo, dejando clara la inclinación terminológica de dichos modelos educativos. Además, con *Glove* y *word2vec* se generó un grafo que ayuda a tener una visión más clara de cada propuesta educativa.

En relación con otros autores, aunque no existe como tal una minería de texto similar educativa, podemos encontrar autores como Leonisio y Strijbis (2012), que apuntaban desde hace una década el aumento de popularidad de hacer minería de texto con *Wordfish* para posiciones ideológicas, dejando ver de manera clara que los modelos educativos 2017 y

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

2022 tienen una clara posición ideológica diferente, por la alternancia de gobiernos. En cuanto a la experimentación de minería de texto en las ciencias sociales, Roberts et al., (2016) reconoce que la construcción de modelado de temas es compleja en cuanto a crear modelos específicos, pero tiene aportaciones generales importantes. Sin duda alguna la inteligencia artificial y aprendizaje de máquina no supervisado y supervisado seguirá avanzando, buscando un aprendizaje profundo de los textos.

El modelo de investigación propuesto puede apoyar a futuros análisis de datos con la minería de datos educativos, y no solamente en comparación de documentos educativos. Usar la minería de texto con el PNL para tener una visión apoyada en la inteligencia artificial de la evolución del Sistema Educativo de México, apoyará a la construcción de categorías integrando los análisis de lectura reflexiva. Es por ello que, buscar un punto de neutralidad ideológica, se convierte en una necesidad para analizar una cantidad considerable de documentos o miles de páginas. Con apoyo de la minería de texto se puede obtener una visión alterna mediante la inteligencia artificial y el aprendizaje de máquina no supervisado, para observar regularidades que no están a la simple vista del ojo humano, dejando bien claro que no se puede sustituir el ojo humano por la máquina, pero sí se pueden apoyar para establecer puntos de reflexión alternos a los tradicionales en el campo educativo. Además, que un conjunto de datos más grande puede generar mayor calidad en el análisis con minería de texto (Quasthoff et al., 2014).

No se trata de posicionar a ningún modelo educativo como mejor o peor, sino comprender que en su momento social e histórico cada uno aportó valor. El modelo 2011 dejó clara la importancia de las asignaturas y el surgimiento de las competencias con estándares de aprendizaje. En el 2017 se fortaleció la equidad y la inclusión, y más claridad al trabajo por competencias, con una visión de eficacia. Ahora, el modelo 2022 plantea un reto diferente con una epistemología del sur, donde el aprendizaje parta de la comunidad. El

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

aprendizaje en comunidad parece ser el camino a la educación en los próximos años, ya que varios países de alto nivel educativo, con una visión política diferente, prestan atención a la comunidad, como Singapur, China y Finlandia (Lewis, 2020). Por lo que México deberá poner su propio sello en esta visión de aprendizaje que rescata la comunidad como punto de partida. El reto quizás no se encuentra en la gestión de un programa y curriculum educativo, el reto será cómo cambiar la visión constructivista a la visión humanista, crítica, con una epistemología del sur, para que los docentes logren con apoyo de las autoridades llevar a buen fin el Marco Curricular y Plan de Estudios 2022.

Como trabajo futuro, se plantea analizar otros modelos educativos, así como la versión final que salga en México, con todas las aportaciones que ha dado el magisterio, seguir profundizando en posibilidades que aporte el procesamiento de lenguaje natural, como análisis de ensayos, tesis, revisiones sistemáticas desde la visión de minería de texto.

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

Referencias

Aleem, A., & Gore, M. M. (2020). Educational data mining methods: A survey. *Proceedings - 2020 IEEE 9th International Conference on Communication Systems and Network Technologies, CSNT 2020*, 182–188. https://doi.org/10.1109/CSNT48778.2020.9115734

Anawis, M. (2014). *The Mining: The Next Data Frontier*. https://www.rdworldonline.com/text-mining-the-next-data-frontier/

Andere, E. (2014). Teachers' perspectives on finnish school education: Creating learning environments. In *Teachers' Perspectives on Finnish School Education: Creating Learning Environments*. Springer International Publishing. https://doi.org/10.1007/978-3-319-02824-8

Chandra, Y. (2020). Online education during COVID-19: perception of academic stress and emotional intelligence coping strategies among college students. *Asian Education and Development Studies*. https://doi.org/10.1108/AEDS-05-2020-0097

Cheng, M. Y., Kusoemo, D., & Gosno, R. A. (2020). Text mining-based construction site accident classification using hybrid supervised machine learning. *Automation in Construction*, *118*, 103265. https://doi.org/10.1016/J.AUTCON.2020.103265

Clark, M. (2018, September 9). *An Introduction to Text Processing and Analysis with R*. https://bit.ly/3yQGpdy

Cowen, R. (2018). Narrating and Relating Educational Reform and Comparative Education. *Educational Governance Research*, *7*, 23–39. https://doi.org/10.1007/978-3-319-61971-2_2

Finnish National Agency for Education. (2016). National core curriculum for basic education 2014: national core curriculum for basic education intended for pupils subject to compulsory education. En *Publications 2016*. Next Print Oy.

Fitz J, Hafid T. (2007). Perspectives on the Privatization of Public Schooling in England and Wales. *Educational Policy*. 21(1):273-296. https://doi/10.1177/0895904806297193

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

Flores, A. (2017). La reforma educativa de México y su Nuevo Modelo Educativo - Dialnet. *Revista Legislativa de Estudios Sociales y de Opinión Pública*, *10*(19), 97–129. https://bit.ly/3KRSKAO

Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, *21*(3), 267–297. https://doi.org/10.1093/PAN/MPS028

Hevia, F. J., Vergara-Lope, S., Velásquez-Durán, A., & Calderón, D. (2022). Estimation of the fundamental learning loss and learning poverty related to COVID-19 pandemic in Mexico. *International Journal of Educational Development*, *88*, 102515. https://doi.org/10.1016/J.IJEDUDEV.2021.102515

INEGI. (2021). *Encuesta para la Medición del Impacto COVID-19 en la Educación (ECOVID-ED) 2020*. https://bit.ly/38ccqIG

Jelodar, H., Wang, Y., Rabbani, M., Xiao, G., & Zhao, R. (2020). A Collaborative Framework Based for Semantic Patients-Behavior Analysis and Highlight Topics Discovery of Alcoholic Beverages in Online Healthcare Forums. *Journal of Medical Systems 2020 44:5*, *44*(5), 1–8. https://doi.org/10.1007/S10916-020-01547-0

Jiménez, N. (2021, July 18). *La Jornada - Privatizar la educación, motivo para juzgar a neoliberales: AMLO*. https://www.jornada.com.mx/notas/2021/07/18/politica/privatizar-la-educacion-motivo-para-juzgar-a-neoliberales-amlo/

Klees, S. J. (2008). A quarter century of neoliberal thinking in education: misleading analyses and failed policies. *Globalisation, Societies and Education*, *6*(4), 311–348. https://doi.org/10.1080/14767720802506672

König, J., Jäger-Biela, D. J., & Glutsch, N. (2020). Adapting to online teaching during COVID-19 school closure: teacher education and teacher competence effects among early career teachers in Germany. *European Journal of Teacher Education*, *43*(4), 608–622. https://doi.org/10.1080/02619768.2020.1809650

Leonisio, R., & Strijbis, O. (2012). El problema de la traducción en el análisis cuantitativo de textos.

Aplicación de Wordscores y Wordfish a las mociones de censura contra el lehendakari Ibarretxe. *Revista Española de Ciencia Política*, *30*, 111–120. https://bit.ly/3agW1Np

Lewis, S. (2020). New Evidence: Governing Schooling Through 'What Works.' *PISA, Policy and the OECD*, 133–170. https://doi.org/10.1007/978-981-15-8285-1_6

Lim, L., & Tan, M. (2018). Culture, pedagogy and equity in a meritocratic education system: Teachers' work and the politics of culture in Singapore, *48*(2), 184–202. https://doi.org/10.1080/03626784.2018.1435974

Lo, S. H., & Hung, C. F. S. (2022). *The politics of education reform in China's Hong Kong*. https://bit.ly/3lAfY4d

López Obrador, A. (2020). Periodo neoliberal en México fue sinónimo de corrupción - YouTube. In *Periodo neoliberal en México fue sinónimo de corrupción*. https://www.youtube.com/watch?v=oq6oI--RBJU

Maamuujav, U. (2021). Examining lexical features and academic vocabulary use in adolescent L2 students' text-based analytical essays. *Assessing Writing*, *49*, 100540. https://doi.org/10.1016/J.ASW.2021.100540

Martínez Iñiguez, J. E., Tobón, S., Serna Huesca, O., & Gómez González, J. A. (2020). Autonomía curricular en educación básica. Una propuesta de innovación en el Modelo Educativo 2017 en México. *Páginas de Educación*, *13*(1), 107–125. https://doi.org/10.22235/pe.v13i1.1914

Maulud, D. H., Zeebaree, S. R. M., Jacksi, K., Sadeeq, M. A. M., & Sharif, K. H. (2021). State of Art for Semantic Analysis of Natural Language Processing. *Qubahan Academic Journal*, *1*(2), 21–28. https://doi.org/10.48161/QAJ.V1N2A44

Moawad, R. A. (2020). Online Learning during the COVID- 19 Pandemic and Academic Stress in University Students. *Revista Romaneasca Pentru Educatie Multidimensionala*, *12*(1Sup2), 100–107. https://doi.org/10.18662/rrem/12.1sup2/252

Ng, P. T. (2020). The Paradoxes of Student Well-being in Singapore, *3*(3), 437–451.

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

https://doi.org/10.1177/2096531120935127

OECD. (2018). *Publications - PISA*. PISA 2018 Results. https://bit.ly/3G9RpUY

ONU. (2013). *Educación - Desarrollo Sostenible*. Objetivo 4: Garantizar Una Educación Inclusiva, Equitativa y de Calidad y Promover Oportunidades de Aprendizaje durante toda la vida para todos. https://bit.ly/3wKS6QF

Pennington, J., Socher, R., & Mnning, C. (2014). *GloVe: Global Vectors for Word Representation*. https://stanford.io/3NmRk2K

Quasthoff, U., Goldhahn, D., & Eckart, T. (2014). *Building Large Resources for Text Mining: The Leipzig Corpora Collection*. 3–24. https://doi.org/10.1007/978-3-319-12655-5_1

Rieble-Aubourg, S., & Viteri, A. (2020). *Hablemos de política educativa en América Latina y el Caribe #1: Educación más allá del COVID-19*. https://doi.org/10.18235/0002654

Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A Model of Text for Experimentation in the Social Sciences, *111*(515), 988–1003. https://doi.org/10.1080/01621459.2016.1141684

Sakurai, S. (2012). Theory and Applications for Advanced Text Mining. *Theory and Applications for Advanced Text Mining*. https://doi.org/10.5772/3115

SEP. (2022a). *Consulta sobre el Plan y programas de estudio 2022.* https://bit.ly/3LCh0Hw

SEP. (2022b, January). *Plan y Programas de Estudio*. https://bit.ly/39EICOS

Silge, J., & Robinson, D. C. N.-Q. 9. D. S. 2017. (2017). *Text mining with R: a tidy approach* (First edit). O'Reilly.

Slapin, J. B., & Proksch, S. O. (2008). A Scaling Model for Estimating Time-Series Party Positions from Texts. *American Journal of Political Science*, *52*(3), 705–722. https://doi.org/10.1111/J.1540-5907.2008.00338.X

Stein-Sparvieri, E. (2010). Text mining e inferencia de defensas en el análisis del discurso en

Sociedad de Investigación
sobre Estudios Digitales S. C.
ISSN: 2683-328X

psicología. *Subjetividad y procesos cognitivos 14*(2), pp.304-313.

Taulé, M., Peris, A., & Rodríguez, H. (2016). Iarg-AnCora: Spanish corpus annotated with implicit arguments. *Language Resources and Evaluation*, *50*(3), 549–584. https://doi.org/10.1007/S10579-015-9334-3

Trstenjak, B., Mikac, S., & Donko, D. (2014). KNN with TF-IDF based Framework for Text Categorization. *Procedia Engineering*, *69*, 1356–1364. https://doi.org/10.1016/J.PROENG.2014.03.129

Vallejo, G. (2021, July). *AMLO: "Se debería juzgar a expresidentes por privatizar la educación."* https://politica.expansion.mx/presidencia/2021/07/17/amlo-se-deberia-juzgar-a-expresidentes-por-privatizar-la-educacion

Verger, A., Fontdevila, C., & Zancajo, A. (2016). The privatization of education: a political economy of global education reform. En *International perspectives on education reform*. Teachers College Press.

Winne, P. H., & Baker, R. S. J. d. (2013). The Potentials of Educational Data Mining for Researching Metacognition, Motivation and Self-Regulated Learning. *Journal of Educational Data Mining*, *5*(1), 1–8. https://doi.org/10.5281/ZENODO.3554619

Zanini, N., & Dhawan, V. (2015). Text Mining: An introduction to theory and some applications. *Research Matters: A Cambridge Assessment Publication*. https://www.cambridgeassessment.org.uk/Images/466185-text-mining-an-introduction-to-theory-and-some-applications-.pdf